

Agentic Workflows II: The Clinical Brain

Study Guide

by Richard J. Young, Ph.D.

richard@deepneuro.ai

Chapter 20 of "Healthcare Analytics and AI: Building Systems That Actually Work"

The essentials: what to know, why it matters at the bedside, and how models fail.

One-sentence thesis. A clinical agent earns its place only when a human still reviews and signs every draft it produces, and the auditor loop, not the model, is what keeps AI-assisted documentation from quietly becoming AI-authored liability.

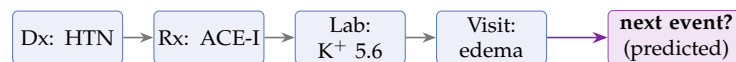
PART 1: The Clinical Brain at Population Scale

Foundation Models on the Whole Library

Earlier chapters built one model per question. This one flips the paradigm. Epic's Cosmos is a de-identified database of more than **300 million** patient records and tens of billions of clinical events, and Epic's Curiosity foundation models, trained on it, read a patient's sequence of medical events, diagnoses, meds, labs, procedures, and predict the most probable next event, exactly the way a language model predicts the next word.

Formally, the model estimates the probability of the next event given everything that came before it:

$$P(e_{t+1} \mid e_1, e_2, \dots, e_t)$$



$$P(e_{t+1} \mid e_1, \dots, e_t) \text{ learned across 300M patient records}$$

A Curiosity-style foundation model reads a patient's ordered event sequence and predicts the most probable next event, exactly as a language model predicts the next word — learning temporal patterns no expert hand-coded, at the cost of a rule you can no longer audit.

ANALOGY

Every clinician works from a personal casebook of a few thousand patients they have seen. Cosmos is the entire library, and Curiosity is a reader who has memorized all of it and guesses your next chapter from how similar stories went.

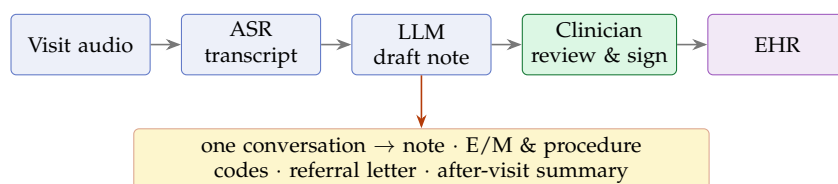
The payoff is that these models can match single-purpose models on length of stay, 30-day readmission, and disease-progression risk without a human hand-coding a single rule, because they learn temporal patterns too subtle or multivariate for any expert to write down. **Watch out.** This is the opposite of the if-then clinical decision support of earlier chapters: the gain is pattern richness, the cost is opacity. You cannot interrogate a foundation model the way you audit a rule, and Epic's data-plus-distribution advantage means the model that quietly shapes care may be one no outside party can independently validate.

When a Scribe Becomes an Agent

An ambient scribe starts as a convenience and quietly becomes an agent. It captures the visit with a microphone, transcribes it, and an LLM turns the transcript into a structured, specialty-specific note that the physician reviews, edits, and signs. Adoption is already large: Dragon Copilot (renamed from DAX in March 2025) is used daily by more than **100,000 clinicians** across 600+ health systems, and Abridge reports more than **250** enterprise health-system partners as of May 2026, up from roughly 100 at the start of 2025, with single deployments as large as 12,000 clinicians at UPMC and a \$5.3 billion valuation after its June 2025 Series E. It is not a tape recorder but a junior

documentation team, turning one conversation into a note, E/M and procedure codes, a referral letter, and a patient-facing after-visit summary, each tailored to a different reader. Because documentation burden is measurable, costly, and immediate, adoption is fast, and when the transcripts also flow into an analytics layer (Microsoft Fabric, Epic's platform), ambient capture becomes a structured-data source for population health, not just a note.

Watch out. The scribe is the first link in the billing chain. A hallucinated detail that inflates code severity is no longer a documentation error; it is a False Claims Act exposure.



The ambient scribe is not a tape recorder: one captured conversation fans out into several tailored documents. It is also the first link in the billing chain, so a hallucinated detail that inflates code severity becomes a False Claims Act exposure — which is why the clinician still reviews and signs.

Consent Before the AI Can Listen

Before any of this runs, someone has to be allowed to press record, and that is a different legal question from privacy. Recording the physician-patient conversation triggers state wiretapping law, which is independent of HIPAA; some states (California, Florida, Illinois, Washington, Pennsylvania) require all-party consent, and a violation there is a criminal-statute problem, not a policy problem.

ANALOGY

HIPAA is the lock on the filing cabinet. Wiretapping law governs whether you were allowed to press “record” in the first place. Two different doors, two different keys.

A system can be fully HIPAA-compliant and simultaneously breaking a state eavesdropping law, and the edge cases, adolescent confidential services, cognitive impairment, unconscious trauma patients, are governed by institutional policy rather than settled law. **Watch out.** Audio is a biometric identifier: a recording of a patient describing symptoms in their own voice is far more re-identifiable than a structured note. Define retention, access, and deletion before deployment, not after the first complaint, because a consent checkbox is not a legal defense.

Epic's Personified Agents

Epic gives its agents names and scopes their autonomy one role at a time, and the base is already broad: as of early 2026 more than **85 percent** of Epic customers were running Epic AI in some form. Art, clinician-facing, generates Insights chart summaries now used over **16 million times per month** and does ambient documentation natively. Emmie, patient-

facing, lives in MyChart. Penny, revenue cycle, automates professional coding with reported 20-percent-plus denial reductions. Agent Factory lets health systems build their own no-code agents. What makes this more than branding is that native beats overlay: an agent that generates a note, stages orders, checks formulary, triggers prior auth, and writes back to the problem list, all inside the EHR's data model, is transformative in a way a pasted note never is.

Agent	Audience	Core capability	Reported adoption
Art	Clinician	Insights summary; chart search; ambient docs	16M monthly Insights uses
Emmie	Patient	MyChart assistant; autonomous follow-up	Live in all 50 states
Penny	Revenue cycle	Pro-fee coding; autonomous ED/radiology	200+ orgs; >20% fewer denials
Agent Factory	Developer	No-code custom agents on Epic FHIR	Early-adopter systems

Order-Entry Agents and Write-Back

Consider a routine sentence: “start lisinopril 10 mg daily and order a BMP in two weeks.” An order-entry agent extracts the order, maps “lisinopril 10 mg” to the right RxNorm code and “BMP” to the right LOINC panel, and stages a FHIR Medication-Request or ServiceRequest for signature, acting as a translator between the language of the exam room and the language of the EHR. It drafts the paperwork; the physician still signs it. This is also a natural insertion point for AI deprescribing: over **40 percent** of patients past 65 take five or more medications, and polypharmacy drives adverse drug events, falls, and

ER visits, so when the agent stages a new script, a deprescribing layer can surface candidates to stop, grounded in the patient's diagnoses and labs.

Watch out. The human-in-the-loop is the whole safety model. An agent that executes an order without physician confirmation is, under current law, practicing medicine without a license, and because these agents ingest free text (transcripts, chart notes), a malicious or garbled string in the source is a prompt-injection vector at the bedside. Treat any tool call the agent proposes as untrusted until a clinician approves it.

PART 4: Verification and the Clinical Auditor Loop

Generated Is Not Verified

A generated note is not a verified note. Ambient notes carry roughly a **1 to 3 percent** error rate, but the fail-

ure modes are distinctive and clinical, not cosmetic, which is why the fix is borrowed from radiology: a second reader, cheap to add, exists precisely because the first read is fallible in ways the author cannot see.

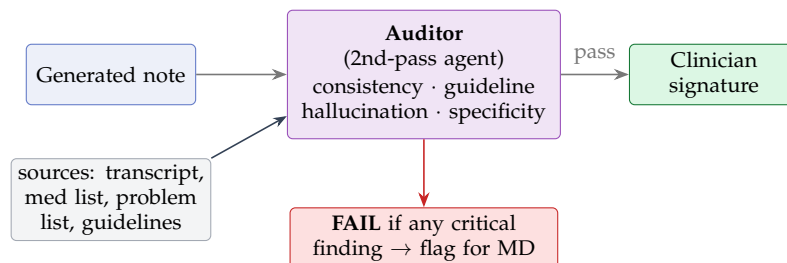
Failure mode	What it is	Example	Detectability
Hallucination	Fabricated but plausible detail	"Patient denies chest pain" (never discussed)	Hard: needs source grounding
Omission	Documented fact missing	Knee pain mentioned, absent from note	Hardest: no negative signal
Misattribution	Right fact, wrong context	Symptom mapped to wrong system	Medium: guideline checks help
Tone drift	Style differs from clinician	Overly formal phrasing	Low priority; easy to spot

Building the Auditor

The second reader here is a second-pass agent, the auditor, sitting between the generated note and the signature. It ingests the note plus its verification sources, the raw transcript, medication list, problem list, and guidelines, and runs typed checks; the decision rule is deliberately blunt, any single critical finding fails the audit and routes the note back to the clinician.

That one rule is what preserves the line between AI-assisted and AI-authored documentation, and in production the whole pipeline adds under five seconds.

Watch out. No technical safeguard replaces physician review. The note bears the physician's signature, and the signature carries the liability, so build the auditor before the scribe reaches production.



The clinical auditor is a second-pass agent between the generated note and the signature. It checks the note against source material with typed checks; any single critical finding fails the audit and routes the note back to the clinician.

Check	Compares	Question	Failure action
Consistency	Note vs transcript + chart	Does every mentioned med appear?	Flag omission
Guideline	Note vs practice guideline	Does new HF dx include echo, BNP, BMP?	Flag gap for MD
Hallucination	Note vs transcript + chart	Can each fact trace to a source?	Fail audit (critical)
Specificity	Note vs ICD-10/CPT rules	Coded to required specificity?	Flag for coder

CMS Signature Guidance

The accountability question has a clear answer. CMS clarified that when a provider uses a scribe, including an AI scribe, the provider must sign the entry to authenticate it: the AI does not sign, the clinician does. A signed AI note supports billing only if it accurately reflects the encounter and meets the same standards as a human-written note, which means systems need a chain of custody, the source material, the model version, the draft before editing, what the clinician changed, when it was signed, and what was billed. **Watch out.** Vendors must treat audit logging, version control, and edit history as product requirements, not afterthoughts; without that record you cannot explain how an AI sentence entered the chart or why a code was supported.

The Productivity Paradox

Saving time does not automatically return it. When ambient AI saves a clinician roughly **35 minutes a day**, the organization faces a choice: give the time back as longer visits and earlier departures, or monetize it as more volume, and under fee-for-service the incentive points at more patients, so burnout returns to baseline. A faster highway does not shorten the commute; it fills with more cars, and AI that removes documentation friction can simply be back-filled with more encounters. Goodson and colleagues (2025) call the “documentation burden equals burnout” framing pre-

mature certainty; a three-million-worker shortage is real, so more throughput may be the only way to preserve access, but throughput without fixing staffing and incentives treats the symptom, not the disease.

Watch out. The same analytics pipeline that creates the productivity gain can measure whether the organization is consuming it responsibly, shrinking visit lengths, rising turnover, declining note quality, or recklessly. Make the trade-off visible.

Should You Build This?

Build it when documentation burden is the problem you are actually solving and the physician remains the reviewer of record: the multi-site burnout reduction (51.9 percent to 38.8 percent at 30 days) came from drafts clinicians read, corrected, and signed. **Think twice when** patient consent for ambient recording is an afterthought, or when the recovered minutes will be monetized as added patient volume, the productivity paradox that returns burnout to baseline. A 1-to-3 percent error rate sounds small until the error is a hallucinated “denies chest pain” that no one catches before it propagates through the chart. **Who bears the risk.** The note bears the physician’s signature and the physician bears the liability; no vendor contract changes that. Patients carry the residue of undetected omissions, and a clinician who stops reading drafts has quietly converted a review tool into an unreviewed author.

Quick Self-Test

1. How does a Curiosity-style foundation model differ, in paradigm, from rule-based clinical decision support, and what does it trade away?
2. A health system is fully HIPAA-compliant and records visits with an ambient scribe in California without patient consent. What law might it still be breaking, and why is that a separate question?
3. Name the three most dangerous ambient-scribe error modes and say which is hardest to detect and why.
4. What single condition makes the clinical auditor return FAIL, and what sources does it check a note against?
5. Under CMS signature guidance, who is accountable for an AI-drafted note, and what chain-of-custody records should exist?
6. State the productivity paradox in one sentence and name one metric that reveals which way an organization went.

If You Remember One Thing

The model drafts; the clinician signs; the clinician and the organization are accountable. Every capability in this chapter (population-scale prediction, ambient scribes, named agents, order entry) collapses back to that line. The auditor loop and the human signature are what hold it. If a hallucinated note inflates a code or omits a material finding, the consequences are clinical, legal, and financial, so build the auditor before the scribe ships.

Richard J. Young, Ph.D. · richard@deepneuro.ai · Senior AI Research Scientist, UnitedHealth Group · UNLV Lee Business School · ORCID 0000-0002-1109-7552